Journal of Combinatorial Theory, Series A **99**, 307–344 (2002)
doi:10.1006/jcta.2002.3273

The Tennis Ball Problem

D. Merlini[1], R. Sprugnoli, and M. C. Verri

Dipartimento di Sistemi e Informatica, via Lombroso 6/17, 50134, Firenze, Italy
E-mail: merlini@dsi.unifi.it, sprugnoli@dsi.unifi.it, verri@dsi.unifi.it

Received February 10, 2000

Mallows and Shapiro, (*J. Integer Sequences* **2** (1999)) have recently considered what they dubbed *the problem of balls on the lawn*. Our object is to explore a natural generalization, *the s-tennis ball problem*, which reduces to that considered by Mallows and Shapiro in the case $s = 2$. We show how this generalization is connected with s-ary trees, and employ the notion of generating trees to obtain a solution expressed in terms of generating functions. © 2002 Elsevier Science (USA)

Key Words: s-ary trees; generating trees; generating functions.

1. INTRODUCTION

Grimaldi and Moser [3] drew attention to the *finite* combinatorial aspects of *the tennis ball problem* introduced by Tymoczko and Henle in their introductory logic text [10, pp. 304–305], where, however, the interest was rather in implementing a process indefinitely; in particular, they demonstrated that this problem yields yet another instance of the Catalan numbers. The Catalan numbers, C_n, $n \geqslant 0$, given by

$$C_n = \frac{1}{n+1}\binom{2n}{n}, \qquad n \geqslant 0,$$

have, of course, proved an immensely popular topic of late. So, it was only to be expected that these further appearance of the sequence would elicit the attention of other researchers. The problem was quickly taken up in greater detail by Mallows and Shapiro, who, in [4], introduced their own slant on it in the following words:

> The tennis ball problem goes as follows. At the first turn, you are given balls numbered one and two. You throw one of them out of the window onto the lawn. At the second turn, balls numbered three and

[1] To whom correspondence should be addressed.

0097-3165/02 $35.00

four are brought in, and now you throw out on the lawn any of the three balls in the room with you. Then balls five and six are brought in, and you throw out one of the four available balls. The game continues for n-turns. The first question is how many different arrangements are possible. [By *different arrangements* here is meant the various sets of balls that might be on the lawn after the n-th turn, irrespective of the order of their arrival on the lawn.] It is easy to see that there are 2, 5, and 14 possibilities after 1, 2, and 3 turns. This suggests a Catalan number, which turns out to be the case. [This result already appears in [3].] A more delicate question is: "What is the total sum [say, Σ] of the balls on the lawn over all these possibilities?" Here, the first few terms are 3, 23, 131, 664.

It is this latter question that Mallows and Shapiro proceed to tackle in [4], their final answer being

$$\Sigma = \Sigma_n^{[2]} = \frac{2n^2 + 5n + 4}{n+2}\binom{2n+1}{n} - 2^{2n+1}.$$

There does not seem to be any direct combinatorial interpretation of this result; and, indeed, in obtaining it, Mallows and Shapiro evince a remarkable facility in manipulating sums of binomial coefficients (after the style of [2]) and, more especially, the generating function for the Catalan numbers, namely

$$C(t) = \sum_{n \geqslant 0} C_n t^n = (1 - \sqrt{1-4t})/2t.$$

Now, it is clear that this *problem of balls on the lawn*, as formulated in [4], admits many natural variants. For a start, attention might be paid to the order in which the balls come out onto the lawn, although we do not do so in this paper. Retaining the regime of unordered sets, we might instead vary the supply of balls, making s balls available at each turn, but continuing to throw out balls one at a time; this is the problem we attack in the present paper and will be called the *s-ball tennis problem*. Then we might go further, being supplied with s balls at each turn, but now throwing out t balls at a time, with $t<s$. This more general problem will not be considered in the present paper; however, we will study the particular case $s=4,\ t=2$.

The following points summarize the main contributions given by our paper:

- The s-tennis ball problem is shown to be equivalent, as a counting problem, to s-ary trees from one side and, from another, to a set of *generating trees*. This allows us to prove a number of facts, simplifying our approach to the s-tennis ball problem and to answer the first question posed by Mallows and Shapiro.

- We find a solution to the sum of the balls in the lawn problem depending on the counting generating function for s-ary trees. In particular, we find that the number of possible arrangements after n turns are counted by $T_{n+1}^{[s]}$, the number of s-ary trees with $n+1$ nodes and give an explicit formula for $\Sigma_n^{[s]}$, the sum of all the balls thrown onto the lawn in n turns, in terms of $T_n^{[s]}$; we also give an explicit formula for $A_n^{[s]}$, a quantity strictly related to $\Sigma_n^{[s]}$ (see Theorem 2.5). These formulas can only be closed in the case $s = 2$, which correspond to the result [4].
- We give an asymptotic approximation for $A_n^{[s]}$.
- Finally, we give a formula for the $(4, 2)$-regime, which is, at the moment, the only case with $t \neq 1$, we have tackled (see the appendix for a sketch of the method used in this case).

The material is organized in the following way. In Section 2, we show the connection between the s-tennis ball problem and a suitable set of generating trees, combinatorial objects whose importance is becoming more and more evident. In Section 3, we study the relations between the s-tennis ball problem, generating trees and s-ary trees. In Section 4, we perform a study of generating functions in order to prove our main result concerning a formula for $A(t) = \sum_n A_n^{[s]} t^n$. In Section 5, we derive from $A(t)$ a series of asymptotic results which solve Mallows and Shapiro's generalized problem. Finally in the appendix, we solve the $(4, 2)$-case.

2. THE TENNIS BALL PROBLEM AND GENERATING TREES

By rephrasing the exposition given in the Introduction, we can say that the s-tennis ball problem goes as follows. Let $s \geqslant 2$ be an integer. At the first turn, you are given balls numbered one through s. You throw one of them out of the window onto the lawn. At the second turn, balls numbered $s+1$ through $2s$ are brought in and now you throw out on the lawn any of the $2s-1$ remaining. Then balls $2s+1$ through $3s$ are brought in and you throw out one of the $3s-2$ available balls. The game continues for n turns. At this point, you pick up the n balls in the lawn and consider the ordered sequence $B = (b_1, b_2, \ldots, b_n)$ with $b_1 < b_2 < \cdots < b_n$. This sequence will be called a *tennis ball s-sequence* and the first question is: how many tennis ball s-sequences of length n exist? The second question is: what is the sum of all the balls in all the possible s-sequences of length n ? Obviously, if we answer to both these questions, we also know the average sum of the balls in an s-sequence of length n. Our first result is a characterization of tennis ball s-sequences:

THEOREM 2.1. *A sequence $B = (b_1, b_2, \ldots, b_n)$ with $1 \leqslant b_1 < b_2 < \cdots < b_n \leqslant sn$ is a tennis ball s-sequence if and only if $b_i \leqslant si$, for every i, $1 < i \leqslant n$.*

Proof. First, let us prove that any sequence with the stated property is a tennis ball s-sequence; in fact, it is sufficient to consider a sequence in which b_i has been thrown at the ith throw. For the converse, let us suppose that $b_i > si$; so b_i has been thrown at the jth throw with $j > i$, when the corresponding balls have been brought in. But in the first i throws, i balls marked $b_r \leqslant si$ should have been thrown, and therefore b_i cannot be in the ith position. ∎

A pictorial way to represent tennis ball s-sequences is by means of a forest of s infinite trees, in which at every level the balls thrown onto the lawn are shown. An s-sequence of length n is just a path in a tree, starting at one of the s roots and extending down for n levels (see Fig. 1). We added a root labeled 1 and transformed the forest into a tree in order to explain a correspondence between such pictorial representation of tennis ball s-sequences and *generating trees*. As far as we know, the concept of generating trees has been used for the first time (without any specific name) by Chung *et al.* [1] to examine the reduced Baxter permutations. This technique has been successively applied to other classes of permutations and the main references on the subject are due to West [11, 12]. Let us give the following:

DEFINITION 2.2. A generating tree is a rooted labeled tree with the property that if v_1 and v_2 are any two nodes with the same label then, for

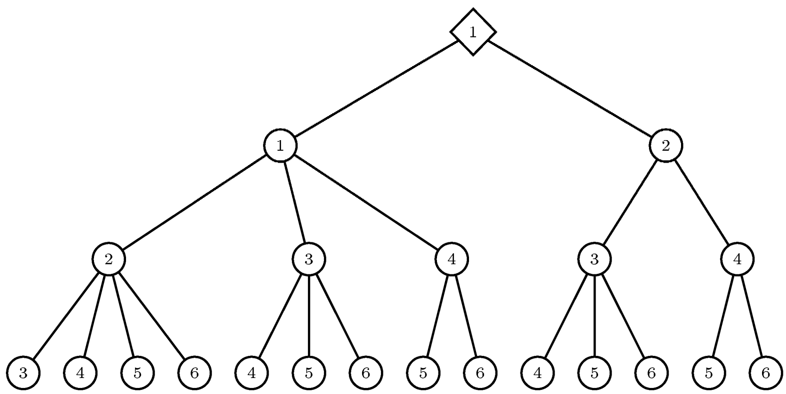

FIG. 1. A pictorial representation of *tennis ball s-sequences* for $s = 2$.

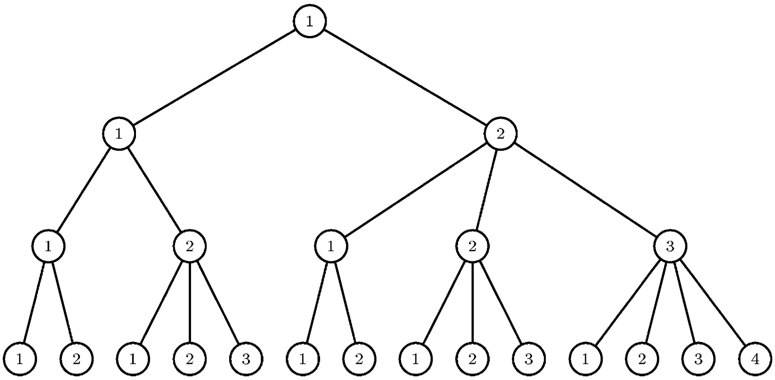

FIG. 2. The partial generating tree for specifications (2.1) and (2.2) when $s = 2$.

each label l, v_1 and v_2 have exactly the same number of children with label l. To specify a generating tree it, therefore, suffices to specify:

(1) the label of the root;

(2) a set of rules explaining how to derive from the label of a parent the labels of all of its children.

We explicitly observe that the tree in Fig. 1 is not a generating tree. On the other hand, Fig. 2 illustrates the upper part of the generating tree which corresponds to the following specification:

$$\begin{cases} \textit{root}: & (1), \\ \textit{rule}: & (k) \to (1) \cdots (k)(k+1). \end{cases} \tag{2.1}$$

The reader can easily observe that the two trees in Figs. 1 and 2 are very similar: in fact, they are isomorphic as we associate a node labeled b at level i (by convention, the level of the root is 0) to a node labeled $2i - b + 1$ in the other tree. In general, if we fix the number s, and b is the label of some node at level i in the tree of the tennis ball s-sequence, let's change it into $si - b + 1$. Its children are labeled $b+1, b+2, \ldots, s(i+1)$ by construction and therefore, by the same transformation, they are changed into $s(i+1) - b - 1 + 1, s(i+1) - b - 2 + 1, \ldots, s(i+1) - s(i+1) + 1 = 1$. In other words, if $k = si - b + 1$ is the label of the node at level i, then its children are labeled $1, 2, \ldots, k + s - 1$, if we read them from right to left. What we get, is therefore the generating tree with specification:

$$\begin{cases} \textit{root}: & (1), \\ \textit{rule}: & (k) \to (1) \cdots (k+s-2)(k+s-1), \qquad s \geqslant 2 \end{cases} \tag{2.2}$$

which corresponds to specification (2.1) when $s = 2$ (Fig. 6 gives the generating tree corresponding to specification (2.2) when $s = 3$). We can, therefore, state the following result:

THEOREM 2.3. *The set of the tennis ball s-sequences of length n is in a* 1–1 *correspondence with the* $(n+1)$*-long paths in the generating tree produced by specification* (2.2).

Proof. Obviously, the introduction of the node at level 0 is irrelevant and simply adds a fixed component 1 at the beginning of the sequence. More precisely, if $(b_1, b_2, \ldots, b_n)$ is a tennis ball s-sequence of length n, then it corresponds to the $(n+1)$-path $(h_0, h_1, \ldots, h_n)$ in the generating tree with $h_0 = 1$ and $1 \leqslant h_i \leqslant h_{i-1} + s - 1,\ 1 \leqslant i \leqslant n.$ ∎

There are two immediate, but important consequences of this theorem. The first one is another proof that the sequences in Mallows and Shapiro's original problem (i.e., tennis ball 2-sequences) are counted by the Catalan numbers:

COROLLARY 2.4. *The set of tennis ball* 2*-sequences of length n are counted by* C_{n+1}, *the* $(n+1)$*th Catalan number.*

Proof. Actually, it is well known (see, e.g., [11]) that the generating tree produced by rule (2.1) is described by the Catalan numbers, in the sense that the number of its nodes at level n is just C_{n+1}. ∎

The second consequence is related to the second question on the sum of the labels in all the s-sequences of length n. It does not answer the question directly, but will be very useful in finding the solution of the problem. Let us denote by $T_n^{[s]} = T_n$ the number of nodes at level $n-1$ in the generating tree; by Corollary 2.4, $T_n^{[2]} = C_n$ and in general, we can prove the second consequence of Theorem 2.3, which we state as a new theorem, due to its importance:

THEOREM 2.5. *If* $S_n^{[s]}$ *is the sum of all the labels at level n in the tree of the tennis ball s-sequences, then* $(sn+1)T_{n+1}^{[s]} - S_n^{[s]}$ *is the sum of all the labels at level n in the corresponding generating tree. Also, if* $\Sigma_n^{[s]}$ *is the sum of all the balls thrown onto the lawn in n turns, then the corresponding sum* $A_n^{[s]}$ *in the generating tree is*

$$A_n^{[s]} = \left(\frac{(sn+2)(n+1)}{2}\right) T_{n+1}^{[s]} - \Sigma_n^{[s]}.$$

Proof. The first part follows immediately from the correspondence $k \to sn + 1 - b$, where b is the label of a node at level n in the tennis ball tree and k the label of the corresponding node in the generating tree. This relation shows that $k + b = sn + 1$; therefore, let us consider the sum of the balls plus the sum of the labels in the generating tree. If we reached level n, we have $T^{[s]}_{n+1}$ possible sequences; the sum at level 0 is therefore $T^{[s]}_{n+1}$; the sum at level 1 is $(s+1)T^{[s]}_{n+1}$; the sum at level 2 is $(2s+1)T^{[s]}_{n+1}$, and so on. The total sum is $(1 + (s+1) + (2s+1) + \cdots + (ns+1))T^{[s]}_{n+1}$ and from this the expression of the theorem follows immediately. ∎

3. THE CORRESPONDENCE WITH *s*-ARY TREES

The correspondence between the generating tree specification (2.2) with $s = 2$ and binary trees has been deeply investigated. In general, the generating tree (2.2) is in a $1 - 1$ correspondence with s-ary trees and this relation is very useful in proving some properties of the generating tree. Therefore, let us give a short introduction to s-ary trees.

DEFINITION 3.1. An *s-ary tree* is either an empty tree or consists in a node (called the *root*) to which an ordered sequence of s (possibly empty) s-ary trees are attached.

A pictorial version of this definition is given in Fig. 3; by applying the Schützenberger methodology (see, e.g., [7]), we immediately obtain a functional equation for the generating function $T(t)$ counting the number T_n of s-ary trees with n nodes:

$$T(t) = 1 + tT(t)^s. \tag{3.3}$$

Except for $s = 2$, when we have $T(t) = (1 - \sqrt{1-4t})/2t$, an explicit formula[2] for $T(t)$ is awkward or is impossible to obtain (for $s \geqslant 5$). However, an expression for T_n, is easily found by means of the Lagrange inversion theorem. By setting $T_n = [t^n]T(t)$ and $w = w(t) = T(t) - 1$, we obtain

$$T_n = [t^n][w+1 \mid w = t(w+1)^s] = \frac{1}{n}[t^{n-1}](t+1)^{sn} = \frac{1}{n}\binom{sn}{n-1}$$
$$= \frac{1}{(s-1)n+1}\binom{sn}{n};$$

[2] We write T instead of $T^{[s]}$ for simplicity sake. All superscripts $[s]$ will be ignored, when they are not strictly necessary to avoid ambiguity.

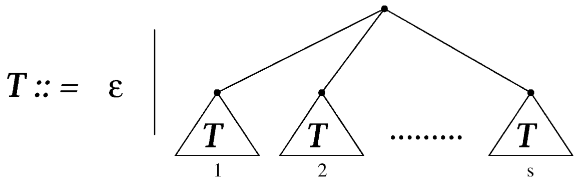

FIG. 3. The definition of s-ary trees.

this is the well-known formula counting s-ary trees; obviously, when $s = 2$, it gives an explicit expression for the Catalan numbers. In order to show the bijection between the generating tree specification (2.2) and s-ary trees we introduce a labeling technique by giving the definition of *e-labeled s-ary trees* (the *e* comes from the term *embedding* below):

DEFINITION 3.2. An e-labeled s-ary tree is an s-ary tree whose nodes are labeled in the following way:

1. if the s-ary tree is complete (that is, no subtrees are empty), then we label the nodes in the following way:

- we label the root by 1;
- recursively, if a node is labeled k, then we label its children by $k, k + 1, \dots, k + s - 1$ proceeding from right to left.

2. if the s-ary tree is not complete, then we label its nodes by embedding it in a complete s-ary tree labeled as in 1).

Since this labeling is unique, the number of e-labeled s-ary trees with n nodes coincides with T_n.

Figure 4 illustrates a complete e-labeled ternary tree. From our point of view, the important result connecting generating trees and e-labeled s-ary trees is the following theorem, which answers Mallows and Shapiro's first question:

THEOREM 3.3. *There exists a* 1–1 *correspondence between the set of e-labeled s-ary trees with* $n + 1$ *nodes and the nodes at level n in generating tree corresponding to specification* (2.2).

Proof. Every node at level n in the generating tree determines a unique $(n + 1)$-path $(h_0, h_1, \dots, h_n)$ in the generating tree; therefore, what we will show is that there is a 1–1 correspondence between e-labeled s-ary trees and paths in the generating tree. Let us define the procedure to pass

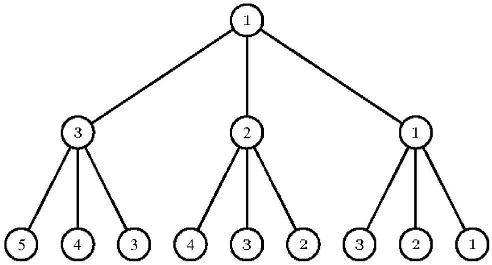

FIG. 4. A complete e-labeled ternary tree.

from an $(n+1)$-path $(h_0, h_1, \ldots, h_n)$ to the corresponding e-labeled s-ary tree:

- The first element h_0 in the $(n+1)$-path is always 1; let it correspond to the root of the tree, labeled 1. Connect the root by means of dotted lines to all its children, labeled 1 through s from right to left;

- After proceeding as far in the path as an element with value h_{i-1}, we will have constructed an e-labeled tree in which some of the edges have only tentative status (denoted by dotted lines instead of solid ones). Now, if h_i is the next element in the path we should have $1 \leqslant h_i \leqslant h_{i-1} + s - 1$: then we change the dotted line arriving at h_i (not necessarily coming from h_{i-1}) into a solid line, delete all the dotted lines on its left and finally connect h_i by means of dotted lines to all its children, labeled h_i through $h_i + s - 1$ from right to left.

- As a final step, when the $(n+1)$-path is exhausted, we eliminate all the dotted lines and the nodes at their ends.

The inverse correspondence is now rather obvious; suppose we have an e-labeled s-ary tree with $n+1$ nodes and simply perform a pre-order visit to the tree. What we get is an $(n+1)$-path in the generating tree because, by construction, its elements satisfy the condition for $(n+1)$-paths, i.e., an element h_{i-1} can only be followed by an h_i such that $1 \leqslant h_i \leqslant h_{i-1} + s - 1$. Obviously, different $(n+1)$-paths correspond to different e-labelled trees, and the proof is complete. ∎

In Fig. 5, we consider the 6-path (1 3 4 3 1 3) (which is a path in the generating tree of Fig. 6) and show the construction of the corresponding e-labeled 3-ary tree. Vice versa, starting with the e-labeled 3-ary tree and visiting it in pre-order, we find the original 6-path.

The generating function $G(t)$ for a generating tree is the formal power series $G(t) = \sum_{k=0}^{\infty} g_k t^k$ in which g_k is the number of its nodes at level k. For our generating trees, $G(t)$ is obtained by observing that the only tree not corresponding to any path in the generating tree is the empty one; this

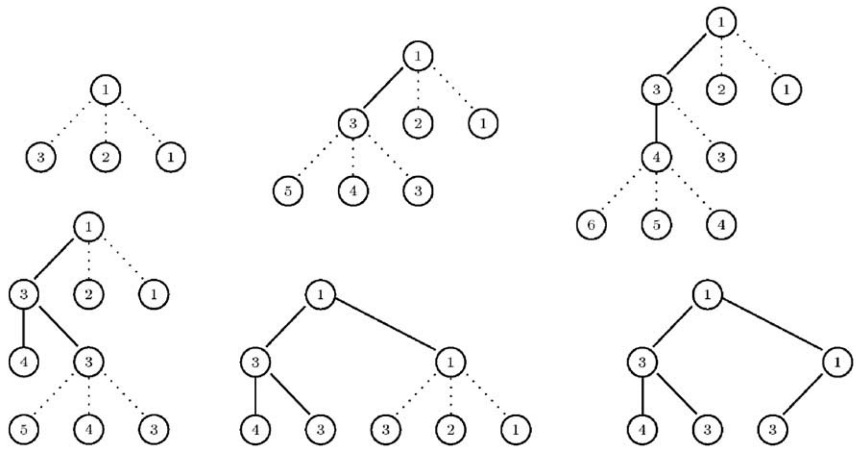

FIG. 5. From the 6-path (1 3 4 3 1 3) to the corresponding e-labeled 3-ary tree.

simply means

$$G(t) = \frac{T(t) - 1}{t} = T(t)^s.$$

We are now going to prove an important result concerning the generating function of the subtrees in the generating tree originating from a node labeled k. Let us call $G_k(t)$ this function, so that we have $G_1(t) = G(t)$, and in general:

THEOREM 3.4. *The generating function $G_k(t)$ for the subtree, whose root is labeled k, is given by*

$$G_k(t) = T(t)^{k+s-1}.$$

Proof. In the correspondence between paths in the generating tree and e-labeled s-ary trees (see Fig. 5), when a node in an e-labeled s-ary tree is labeled k, it is followed by $k + s - 1$ (possibly empty) s-ary trees, i.e., the trees originating from the dotted lines in the proof of Theorem 3.3. This immediately proves the desired relation. ∎

4. THE MAIN RESULT

Up to this moment, we have presented a series of results which will now allow us to obtain our main result, i.e., the answer to Mallows and Shapiro's

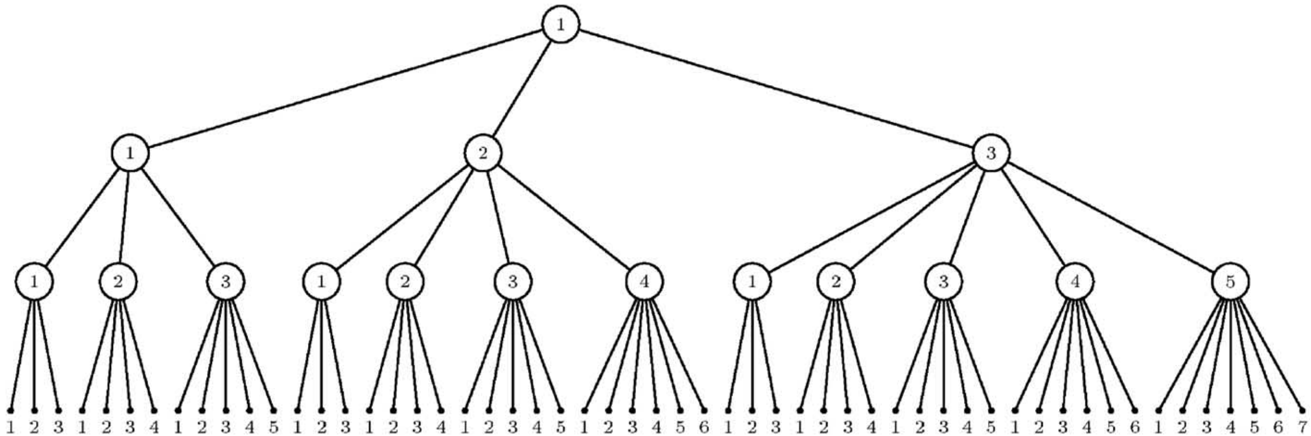

FIG. 6. The partial generating tree for specification (2.2) when $s = 3$.

second question in terms of generating functions. In the next section, we will see how to pass from these generating functions to a series of more specific counting results. The correspondence between tennis ball s-sequences and generating trees allows us to use results obtained for these latter combinatorial objects. First of all, we re-introduce from [6] the concept of an AGT matrix.

DEFINITION 4.1. An infinite matrix $\{d_{n,k}\}_{n,k\in\mathbf{N}}$ is said to be "associated" to a generating tree with root (c) (AGT matrix for short) if $d_{n,k}$ is the number of nodes at level n with label $k+c$. By convention, the level of the root is 0.

Tables 1 and 2 illustrate the AGT matrices associated to the generating tree specification (2.2) with $s=2$ and 3, and corresponding to Figs. 2 and 6, respectively. We explicitly observe that Table I corresponds to the well-known Catalan triangle.

From here on, $D=\{d_{n,k}\}_{n,k\in\mathbf{N}}$ will denote the AGT matrix associated to the generating tree specification (2.2). This matrix plays an important role in our development, and therefore will be deeply studied. In order to explain this role, let us begin by some notations. Let A_n be the total sum of the labels in all the paths from level 0 to level n in the generating tree; we also denote

TABLE 1
The AGT Matrix for Specification (2.2) with $s=2$

n/k	0	1	2	3	4
0	1				
1	1	1			
2	2	2	1		
3	5	5	3	1	
4	14	14	9	4	1

TABLE 2
The AGT Matrix for Specification (2.2) with $s=3$

n/k	0	1	2	3	4	5	6	7	8	9	10
0	1										
1	1	1	1								
2	3	3	3	2	1						
3	12	12	12	9	6	3	1				
4	55	55	55	43	31	19	10	4	1		
5	273	273	273	218	163	108	65	34	15	5	1

by $\alpha_{i,n}$ the sum of the labels at level i in the sum A_n; to compute $\alpha_{i,n}$ we should take into account that a label k at level i is repeated as many times as the number of its successors at level n. Obviously, we have $A_n = \sum_{i=0}^{n} \alpha_{i,n}$ and we have this fundamental result:

THEOREM 4.2. *The sum of the labels at level i in all the paths from* 0 *to n in the generating tree produced by specification* (2.2) *is*

$$\alpha_{i,n} = \sum_{r \geqslant 0} d_{i,r}(r+1)[t^{n-i}]T(t)^{r+s}.$$

Proof. At level i in the generating tree we find labels $k = 1, 2, \ldots, i + s - 1$, so let us interpret r as $k-1$. By definition, $d_{i,r} = d_{i,k-1}$ is the multiplicity of label k at level i; this multiplicity should be multiplied by the "weight" of the label, i.e., $k = r+1$. Finally, as noted before, this quantity is to be considered as many times as the number of successors of the node at level n; by Theorem 3.4 in the previous section, this is counted by the coefficients of $T(t)^{k+s-1} = T(t)^{r+s}$ and, in our case, by the coefficient of t^{n-i}. This concludes our proof. ∎

So, it is now the moment to study the AGT matrix and to this purpose let us denote by $d_r(t)$ the generating function of column r in $D = \{d_{n,r}\}_{n,r \in \mathbf{N}}$. An obvious observation is that $d_{n+1,0}$ coincides with the number of nodes at the previous level n: in fact, $d_{n+1,0}$ denotes the number of labels 1 at level $n+1$, but every node at level n produces one and only one node labeled 1 at the next level. This immediately gives $d_0(t) = T(t)$. The generating function of the other columns is given by the following:

THEOREM 4.3. *The column generating functions of matrix D are*: $d_0(t) = T(t)$, $d_i(t) = T(t) - 1$, *for* $i = 1, \ldots, s-1$, *and*

$$d_r(t) = d_{r-1}(t) - t d_{r-s}(t), \qquad r \geqslant s.$$

Proof. First, we observe that specification (2.2) can be written in the product notation (see [6]): $(k) \to \prod_{j=1}^{k+s-1}(k+s-j)$, $s \geqslant 2$, and this immediately shows that a node at level $n+1$ with label $k+2$ can be determined from the nodes at level n with label $h+1$, such that $h+1+s-j = k+2$, i.e., $h = k-s+1+j$, $j \geqslant 1$. Hence, we have

$$d_{n+1,k+1} = d_{n,k-s+2} + d_{n,k-s+3} + d_{n,k-s+4} + \cdots \tag{4.4}$$

and

$$d_{n+1,k} = d_{n,k-s+1} + d_{n,k-s+2} + d_{n,k-s+3} + \cdots \ .$$

It follows that

$$d_{n+1,k} = d_{n,k-s+1} + d_{n+1,k+1}$$

and finally

$$d_{n+1,k+1} = d_{n+1,k} - d_{n,k-s+1}.$$

By going on to generating functions, we obtain the recurrence relation

$$d_{k+1}(t) = d_k(t) - td_{k+1-s}(t), \qquad k \geqslant s-1.$$

Finally, we observe that relation (4.4) holds true for the number of nodes at level $n+1$ with label $i+1$, $i = 1, \ldots, s-1$, but it cannot begin with negative indices; therefore, in all these cases the result is always the same, i.e., $d_{n+1,i} = d_{n+1,0}$; consequently, the generating functions are all equal to $T(t) - 1$ and the proof is complete. ∎

This theorem allows us to express all the generating functions $d_r(t)$ in terms of $T(t)$:

THEOREM 4.4. *The column generating functions of matrix D can be expressed in terms of $T(t)$ by the following formula*:

$$d_r(t) = p_r(t)T(t) - q_r(t), \qquad r \geqslant 0,$$

where $p_r(t)$ and $q_r(t)$ are polynomials defined by the following relations:

$$p_r(t) = p_{r-1}(t) - tp_{r-s}(t), \qquad q_r(t) = p_{r-1}(t), \quad r \geqslant s$$

with initial conditions:

$$p_i(t) = 1, \quad i = 0, \ldots, s-1, \qquad q_0(t) = 0.$$

Proof. The theorem is obviously true for $r = 0, \ldots, s-1$. Let us suppose it holds true for all $q \leqslant r$ and prove it for $r+1$. From Theorem 4.3, we have

$$d_{r+1}(t) = d_r(t) - td_{r+1-s}(t)$$

and by the induction hypothesis:

$$\begin{aligned} d_{r+1}(t) &= p_r(t)T(t) - q_r(t) - t(p_{r+1-s}(t)T(t) - q_{r+1-s}(t)) \\ &= (p_r(t) - tp_{r+1-s}(t))T(t) - (q_r(t) - tq_{r+1-s}(t)) \\ &= p_{r+1}(t)T(t) - p_r(t). \quad ∎ \end{aligned}$$

TABLE 3
Polynomials $p_r(t)$ for $r \leqslant 8$ and $s = 2, 3, 4$

	s		
	2	3	4
$p_0(t)$	1	1	1
$p_1(t)$	1	1	1
$p_2(t)$	$1-t$	1	1
$p_3(t)$	$1-2t$	$1-t$	1
$p_4(t)$	$1-3t+t^2$	$1-2t$	$1-t$
$p_5(t)$	$1-4t+3t^2$	$1-3t$	$1-2t$
$p_6(t)$	$1-5t+6t^2-t^3$	$1-4t+t^2$	$1-3t$
$p_7(t)$	$1-6t+10t^2-4t^3$	$1-5t+3t^2$	$1-4t$
$p_8(t)$	$1-7t+15t^2-10t^3+t^4$	$1-6t+6t^2$	$1-5t+t^2$

In Table 3, we give the first nine polynomials for $s = 2,\ 3$, and 4, while the following theorem gives a bivariate generating function for the same polynomials:

THEOREM 4.5. *Polynomials $p_r(t)$ can be expressed by the following formula*:

$$p_r(t) = [w^r]\frac{1}{1 - w + tw^s} = [w^r]P(t, w) = [w^r]\sum_{r,k \geqslant 0} p_{r,k}t^k w^r.$$

Proof. From Theorem 4.3 we have a formula valid $\forall r \geqslant 0$:

$$p_{r+s}(t) = p_{r+s-1}(t) - tp_r(t)$$

and by passing to generating functions, we get

$$\frac{1}{w^s}\sum_{r \geqslant 0} p_{r+s}(t)w^{r+s} = \frac{1}{w^{s-1}}\sum_{r \geqslant 0} p_{r+s-1}(t)w^{r+s-1} - t\sum_{r \geqslant 0} p_r(t)w^r.$$

This gives

$$\frac{P(t, w) - p_0(t) - p_1(t)w - \cdots - p_{s-1}w^{s-1}}{w^s}$$

$$= \frac{P(t, w) - p_0(t) - \cdots - p_{s-2}w^{s-2}}{w^{s-1}} - tP(t, w)$$

or

$$P(t, w) - 1 - w - \cdots - w^{s-1} = wP(t, w) - w - \cdots - w^{s-1} - tw^s P(t, w).$$

Hence, we have

$$P(t,w) = \frac{1}{1-w+tw^s}. \quad \blacksquare$$

Finally, we are in a position to prove our main result. If $A(t)$ is the generating function $\sum_{n=0}^{\infty} A_n t^n$, we have:

THEOREM 4.6. *The generating function $A(t)$ of sequence A_n is given by the following formula*:

$$A(t) = (T(t)-1)\frac{2T(t)+s(s-3)(T(t)-1)}{2tT(t)(1-stT(t)^{s-1})^2} = \frac{s(s-1)tT'(t)^2}{2T(t)} + T'(t).$$

Proof. From Theorem 4.2, we have

$$\alpha_{i,n} = \sum_{r\geqslant 0} d_{i,r}(r+1)[t^{n-i}]T(t)^{r+s} = [t^{n-i}]T(t)^s\left(\sum_{r\geqslant 0} d_{i,r}rT(t)^r + \sum_{r\geqslant 0} d_{i,r}T(t)^r\right).$$

Theorems 4.4 and 4.5 yield

$$d_{i,r} = [w^i](p_r(w)T(w) - p_{r-1}(w)),$$

hence

$$\begin{aligned}\sum_{r\geqslant 0} d_{i,r}T(t)^r &= [w^i]\sum_{r\geqslant 0}(p_r(w)T(w)-p_{r-1}(w))T(t)^r\\ &= [w^i](T(w)-T(t))\sum_{r\geqslant 0} p_r(w)T(t)^r\\ &= [w^i](T(w)-T(t))P(w,T(t))\\ &= [w^i]\frac{T(w)-T(t)}{1-T(t)+wT(t)^s}.\end{aligned}$$

The denominator is divisible by $T(w)-T(t)$ and, in fact, by performing a classical polynomial division, we have

$$1-T(t)+wT(t)^s = (T(w)-T(t))\left(1-w\sum_{k=1}^{s} T(t)^{s-k}T(w)^{k-1}\right);$$

hence we obtain

$$\sum_{r\geqslant 0} d_{i,r}T(t)^r = [w^i]\,\frac{1}{1-wV(t,w)},$$

where

$$V(t,w) = \sum_{k=1}^{s} T(t)^{s-k}T(w)^{k-1}.$$

By proceeding in a similar way and by setting

$$Q(t,w) = w\,\frac{\partial P(t,w)}{\partial w} = \sum_{r,k\geqslant 0} rp_{r,k}t^k w^r = \frac{w(1-stw^{s-1})}{(1-w+tw^s)^2}$$

and

$$R(t,w) = \sum_{k=1}^{s}(s-k)T(t)^{s-k-1}T(w)^{k-1},$$

we get

$$\sum_{r\geqslant 0} rd_{i,r}T(t)^r = [w^i]((T(w)-T(t))Q(w,T(t)) - T(t)P(w,T(t)))$$

$$= [w^i]\,\frac{T(t)(T(w)-T(t))(1-swT(t)^{s-1})}{(1-T(t)+wT(t)^s)^2} - [w^i]\frac{T(t)}{1-T(t)+wT(t)^s}$$

$$= [w^i]\,\frac{T(t)(1-swT(t)^{s-1})}{(1-wV(t,w))^2(T(w)-T(t))}$$

$$-\ [w^i]\,\frac{T(t)}{(1-wV(t,w))(T(w)-T(t))}$$

$$= [w^i]\,\frac{wT(t)(V(t,w)-sT(t)^{s-1})}{(1-wV(t,w))^2(T(w)-T(t))} = [w^i]\,\frac{wT(t)R(t,w)}{(1-wV(t,w))^2}.$$

Finally, we have

$$\alpha_{i,n} = [t^{n-i}]T(t)^s[w^i]\left(\frac{wT(t)R(t,w)}{(1-wV(t,w))^2} + \frac{1}{1-wV(t,w)}\right).$$

It is now a simple matter to see that $A(t)$ can be determined by putting $w = t$ in the previous expression; more precisely

$$\begin{aligned}
A_n = \sum_{i\geqslant 0} \alpha_{i,n} &= [t^n]T(t)^s\left(\frac{tT(t)R(t,t)}{(1-tV(t,t))^2}+\frac{1}{1-tV(t,t)}\right)\\
&= [t^n]T(t)^s\left(\frac{s(s-1)tT(t)^{s-1}}{2(1-stT(t)^{s-1})^2}+\frac{1}{1-stT(t)^{s-1}}\right)\\
&= [t^n]T(t)^s\frac{2+s(s-3)tT(t)^{s-1}}{2(1-stT(t)^{s-1})^2}\\
&= [t^n](T(t)-1)\frac{2T(t)+s(s-3)(T(t)-1)}{2tT(t)(1-stT(t)^{s-1})^2},
\end{aligned}$$

since we have $V(t,t) = sT(t)^{s-1}$ and $R(t,t) = s(s-1)T(t)^{s-2}/2$. Finally, from formula (3.3) we have $T'(t) = T(t)^s/(1-stT(t)^{s-1})$, hence

$$A(t) = \frac{s(s-1)tT'(t)^2}{2T(t)} + T'(t). \quad ∎$$

This is our main result; unfortunately, the expression for $A(t)$ only has simple coefficients when $s = 2$ and the Catalan numbers are involved. When $s > 2$, it is very unlikely that we can find closed formulas. Because of that, the next section is dedicated to find asymptotic values for A_n and, consequently, for all the quantities depending on the A_n's, such as the sum of the balls in the lawn.

5. COUNTING RESULTS

The first result regards function $A(t)$:

COROLLARY 5.1. *Function $A(t)$ can be decomposed in the following way*:

$$A(t) = \frac{1}{t}G(t) + T'(t)$$

with

$$\begin{aligned}
G(t) &= \frac{s(s-1)T(t)(T(t)-1)^2}{2((s-1)T(t)-s)^2}\\
&= \frac{s}{2(s-1)^2}\left(2+(s-1)T(t)+\frac{2s+1}{(s-1)T(t)-s}+\frac{s}{((s-1)T(t)-s)^2}\right).
\end{aligned}$$

Proof. From Theorem 4.6 and by repeated applications of identities $T(t) = 1 + tT(t)^s$, $tT(t)^s = T(t) - 1$ and $T'(t) = T(t)^s/(1 - stT(t)^{s-1})$, we have

$$G(t) = \frac{s(s-1)t^2T'(t)^2}{2T(t)} = \frac{s(s-1)t^2T(t)^{2s}}{2T(t)(1-stT(t)^{s-1})^2} = \frac{s(s-1)T(t)(T(t)-1)^2}{2((s-1)T(t)-s)^2}.$$

The second identity can be obtained by a partial fraction decomposition. ∎

From this expression we can find an explicit formula for A_n. Unfortunately, it is not completely closed because of Y_n:

THEOREM 5.2. *Coefficients A_n can be expressed by the following formula*:

$$A_n = \tfrac{1}{2}\, Y_{n+1} - \tfrac{1}{2}((2s-3)n + 2s - 2)T_{n+1},$$

where

$$T_n = \frac{1}{(s-1)n+1}\binom{sn}{n}, \qquad Y_n = \sum_{k=0}^{n}\binom{sk}{k}\binom{s(n-k)}{n-k}.$$

For what concerns the s-tennis ball problem, we have

$$\Sigma_n = \tfrac{1}{2}(sn^2 + (3s-1)n + 2s)T_{n+1} - \tfrac{1}{2}\, Y_{n+1}.$$

Proof. From Corollary 5.1, we have

$$\begin{aligned}
A_n &= [t^{n+1}]G(t) + [t^n]T'(t)\\
&= [t^{n+1}]\,\frac{s}{2(s-1)^2}\Bigg(2 + (s-1)T(t) + \frac{2s+1}{(s-1)T(t)-s}\\
&\quad + \frac{s}{((s-1)T(t)-s)^2}\Bigg) + (n+1)[t^{n+1}]T(t)\\
&= \frac{s}{2(s-1)^2}\Bigg((s-1)T_{n+1} - (2s+1)[t^{n+1}]\frac{1}{s-(s-1)T(t)}\\
&\quad + s[t^{n+1}]\,\frac{1}{(s-(s-1)T(t))^2}\Bigg) + (n+1)T_{n+1}.
\end{aligned}$$

Now, from the Lagrange inversion formula $[t^n][F(w)/(1 - t\phi'(w))|w = t\phi(w)] = [t^n]F(t)\phi(t)^n$, and by setting $w(t) = T(t) - 1$, we have

$$\begin{aligned} X_n &= [t^n]\frac{1}{s-(s-1)T(t)} = [t^n]\left[\frac{1}{1+(1-s)w}\middle| w = t(1+w)^s\right] \\ &= [t^n]\left[\frac{1}{1+w}\frac{1}{1-st(1+w)^{s-1}}\middle| w = t(1+w)^s\right] \\ &= [t^n]\frac{1}{1+t}(1+t)^{sn} = \binom{sn-1}{n}. \end{aligned}$$

Moreover, by setting $Y_n = \sum_{k=0}^{n}\binom{sk}{k}\binom{s(n-k)}{n-k}$, we have, for $n > 0$,

$$Z_n = [t^n]\frac{1}{((s-1)T(t)-s)^2} = \sum_{k=0}^{n} X_k X_{n-k} = \frac{(s-1)^2}{s^2}\,Y_n + \frac{2}{s}X_n,$$

and sum Y_n as far as we know, can be closed only when $s = 2$ (see Theorem 5.3). Finally, putting all together, we get

$$A_n = \frac{s}{2(s-1)^2}\left((s-1)T_{n+1} - (2s+1)X_{n+1} + sZ_{n+1}\right) + (n+1)T_{n+1}$$

and after some simplifying, using in particular the relation $X_n = \frac{(s-1)((s-1)n+1)}{s}T_n$, valid for $n > 0$, we obtain the formula in the statement of the theorem. The formula for Σ_n follows directly from Theorem 2.5. ∎

In Table 4 we give some values for A_n and Σ_n. The case $s = 2$ was already solved by Mallows and Shapiro [4], following a completely different approach. We find again their result, just to show that our development is correct, at least in that case:

TABLE 4
Some Values for A_n and Σ_n

n	0	1	2	3	4	5	6	7
$A_n^{[2]}$	1	5	22	93	386	1586	6476	26 333
$\Sigma_n^{[2]}$	0	3	23	131	664	3166	14 545	65 187
$A_n^{[3]}$	1	9	69	502	3564	24 960	173 325	1 196 748
$\Sigma_n^{[3]}$	0	6	75	708	5991	47 868	369 315	2 783 448

THEOREM 5.3. *For $s = 2$, we have*

$$A_n = \frac{1}{2}\left(4^{n+1} - \binom{2n+2}{n+1}\right), \qquad \Sigma_n = \frac{2n^2+5n+4}{n+2}\binom{2n+1}{n} - 2^{2n+1}.$$

Proof. When $s = 2$, we have

$$Y_n^{[2]} = \sum_{k=0}^{n} \binom{2k}{k}\binom{2(n-k)}{n-k} = [t^n]\frac{1}{1-4t} = 4^n$$

and the proof follows directly from Theorem 5.2. ∎

We now find asymptotic approximations for the coefficients Z_n and A_n as defined in Theorem 5.2:

COROLLARY 5.4. *The coefficients Z_n can be approximated by the following formula*:

$$Z_n \approx \frac{s^{sn-1}}{2(s-1)^{(s-1)n-1}}\left(1 + \frac{s+1}{3\sqrt{2s(s-1)}4^{n-1}}\binom{2n}{n}\right).$$

Proof. By setting $w = w(t) = T(t) - 1$, we have

$$\begin{aligned} Z_n &= [t^n]\frac{1}{(s-(s-1)T(t))^2} = [t^n]\left[\frac{1}{(1+(1-s)w)^2}\middle| w = t(w+1)^s\right] \\ &= [t^n]\left[\frac{1}{(w+1)(1+(1-s)w)}\frac{1}{1-st(w+1)^{s-1}}\middle| w = t(w+1)^s\right]. \end{aligned}$$

We can find an asymptotic approximation of Y_n by using the results in [9] which can be applied to a function $F(w)/(1 - t\phi'(w))$ with $w = t\phi(w)$ and $w(0) = 0$ when $w = w(t)$ only has a finite number of singularities on the circle of convergence and $F(w)$ and $\phi(w)$ "behave well" in that circle. In our case we have $F(w) = \frac{1}{(w+1)(1+(1-s)w)}$ and $\phi(w) = (w+1)^s$, that is,

$$Z_n = [t^n]\left[\frac{F(w)}{1-t\phi'(w)}\middle| w = t\phi(w)\right].$$

The singularities of $w = w(t)$ are among the solutions of the system:

$$\begin{cases} \mathscr{F}(t, w) = w - t\phi(w) = 0, \\ \mathscr{F}'_w(t, w) = 1 - t\phi'(w) = 0 \end{cases}$$

and in this case, we find the solution

$$(r, \gamma) = \left(\frac{1}{s-1} \left(\frac{s-1}{s} \right)^s, \frac{1}{s-1} \right).$$

The function $F(w)$ has singularities $w = -1$ and $1/(s-1) = \gamma$, hence the dominating singularity of $F(w)/(1 - t\phi'(w))$ is (r, γ). We proceed by looking for an asymptotic development of $w = w(t)$ around $t = r$. In a neighborhood of (r, γ), $w(t)$ can be expanded into a series as follows:

$$w(t) = \gamma + a_1 \left(1 - \frac{t}{r}\right)^{1/2} + a_2 \left(1 - \frac{t}{r}\right) + a_3 \left(1 - \frac{t}{r}\right)^{3/2} + \cdots$$

and the values of $a_1, a_2, a_3, \ldots$ can be determined by feeding this expression into $\mathscr{F}(t, w)$, by performing computations and equating coefficients to 0 in order to satisfy equation $\mathscr{F}(t, w) = 0$. In this case, we stopped at the second term and found

$$a_1 = -\frac{\sqrt{2s(s-1)}}{(s-1)^2}, \qquad a_2 = \frac{2(s+1)}{3(s-1)^2},$$

hence, by setting $A = (1 - t/r)$, we have

$$w(t) = \frac{1}{s-1} - \frac{\sqrt{2s(s-1)}\sqrt{A}}{(s-1)^2} + \frac{2(s+1)A}{3(s-1)^2} + O(A^{3/2}).$$

Now, we can use this expression and a system for symbolic computation like Maple to find the following series development:

$$Z_n = [t^n] \frac{F(w)}{1 - t\phi'(w)} = [t^n] \frac{(s-1)}{2sA} + \frac{(s-1)(s+1)\sqrt{2}}{3s\sqrt{s(s-1)}\sqrt{A}} + O(A^{1/2})$$

from which it is now easy to find the asymptotic expression in the statement of the corollary. ∎

Our final result is a formula for A_n; by Theorem 2.5 this immediately gives an asymptotic formula for Σ_n.

THEOREM 5.5. *Coefficients A_n can be approximated by the following formula*:

$$A_n \approx \frac{s^{1+s+sn}}{4(s-1)^{s+(s-1)n}}\left[1 - \frac{(4s-5)\sqrt{2}}{3\sqrt{s(s-1)}4^n}\binom{2n}{n}\right].$$

Proof. The proof is similar to that of Corollary 5.4. From Theorem 4.6 and by setting $w = w(t) = T(t) - 1$, we have

$$\begin{aligned}
A_n &= [t^n]\,\frac{(T(t)-1)(2T(t)+s(s-3)(T(t)-1))}{2tT(t)(1-stT(t)^{s-1})^2}\\
&= [t^n]\left[\frac{((s-1)(s-2)w+2)(w+1)^s}{2(1+(1-s)w)}\frac{1}{1-st(w+1)^{s-1}}\middle| w = t(w+1)^s\right]\\
&= [t^n]\left[\frac{F(w)}{1-t\phi'(w)}\middle| w = t\phi(w)\right]
\end{aligned}$$

with

$$F(w) = \frac{((s-1)(s-2)w+2)(w+1)^s}{2(1+(1-s)w)}, \qquad \phi(w) = (w+1)^s.$$

At this point we can proceed as in Corollary 5.4. In particular, since we have the same dominating singularity, we can use the same development for $w(t)$ around (r, γ) and finally, by setting $A = (1 - t/r)$, find

$$A_n = [t^n]\,\frac{F(w)}{1-t\phi'(w)} = [t^n]\frac{s(\frac{s}{s-1})^s}{4A} - \frac{s(4s-5)\sqrt{2}(\frac{s}{s-1})^s}{12\sqrt{A}\sqrt{s(s-1)}} + O(A^{1/2})$$

from which we find the formula in the statement of the theorem. ∎

For example, by using Theorem 5.2, we find $A_{100}^{[2]} = 0.3033675579 \times 10^{61}$, $A_{100}^{[3]} = 0.1993705295 \times 10^{84}$, and $A_{100}^{[4]} = 0.1410430456 \times 10^{99}$, while Theorem 5.5 gives $A_{100}^{[2]} \approx 0.3032779059 \times 10^{61}$, $A_{100}^{[3]} \approx 0.1992629917\times 10^{84}$ and $A_{100}^{[4]} \approx 0.1409630260 \times 10^{99}$, and this is not too bad.

6. CONCLUSIONS

We have generalized the tennis ball problem, proposed and solved by Mallows and Shapiro [4] in the case $s = 2$, by considering the possibility of handling s balls ($s \geqslant 2$) at each turn, of which we throw out one. As mentioned in the Introduction, a further generalization should be considering the possibility of handling s balls at each turn, of which t are thrown out. The principal reason for concentrating on the problem with $t = 1$ was that we expected the answer to the question about the number of possible sets on the lawn to be a generalized Catalan number, as turned out to be the case. In further work, we have completely solved the $(4, 2)$-case, for which we find the following asymptotic value for the total sum of $2l$-long sequences:

$$\bar{W}_\ell \sim \frac{2\sqrt{\ell}(3\sqrt{2}-4)16^{\ell+1}}{\sqrt{\pi}} - \frac{2 \times 16^{\ell+1}}{(2+\sqrt{2})^2} = O(\sqrt{\ell}\, 16^{\ell+1}).$$

The solution we found is rather complicated and requires some concepts we have not introduced in the previous sections. Therefore, we have separated it from the rest of our paper and give our solution in the appendix.

APPENDIX A

The $(4, 2)$-case of the tennis ball problem generalizes the original problem in the following sense: we are given 4 balls at each turn and throw on the lawn 2 out of the balls remaining in the room. The process can be visualized as a tree if we use a "normalized" version: imagine we have thrown balls numbered 1, 4 at the first turn and balls numbered 2, 3 at the second. This cannot be distinguished from the case when we have thrown balls 1, 2 at the first turn and balls 3, 4 at the second. Therefore, let us suppose that at each turn we throw the two balls with the smallest available numbers in a given sequence; for example, the launches $(1, 4)$, $(5, 8)$, $(2, 10)$ are normalized to $(1, 2)$, $(4, 5)$, $(8, 10)$. With this convention, the tree begins as in Fig. A.1, where the root $(0, 0)$ is conventional, but, as we shall see, is consistent with the rest of the tree.

Having this in mind, we see that the larger element in every couple is important because it determines the possible couples at the next level. If we denote by $M_m^{[\ell]}$ the number of couples at level ℓ with larger element equal to m (or, equivalently, the number of $(4, 2)$-sequences of 2ℓ elements, having m as the largest one), we have

$$M_2^{[1]} = 1, \qquad M_3^{[1]} = 2, \qquad M_4^{[1]} = 3,$$

$$M_4^{[2]} = 1, \qquad M_5^{[2]} = 4, \qquad M_6^{[2]} = 10, \qquad M_7^{[2]} = 16, \qquad M_8^{[2]} = 22.$$

If we denote by $M_\ell = \sum_m M_m^{[\ell]}$ the total number of sequences with 2ℓ elements, we have $M_1 = 6$, $M_2 = 53$, and we can define $M_0^{[0]} = 1$. We now

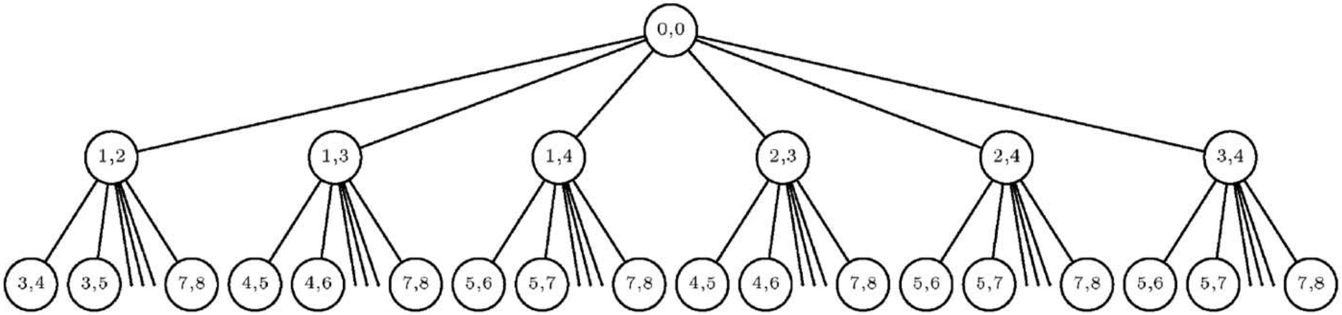

FIG. A.1. The tree for the (4,2)-case.

determine $M_m^{[\ell+1]}$ $(2\ell+2<m\leqslant 4\ell+4)$ by the following observation in which, by a momentary abuse of language, we denote sets with the same symbol as their cardinality: a couple in $M_m^{[\ell+1]}$ is obtained:

- by appending to a couple in $M_{m-2}^{[\ell]}$ the couple $(m-1,m)$;
- by appending to a couple in $M_{m-3}^{[\ell]}$ one of the two couples $(m-1,m)$ and $(m-2,m)$;
- by appending to a couple in $M_{m-4}^{[\ell]}$ one of the three couples $(m-1,m)$, $(m-2,m)$ and $(m-3,m)$;

and so on. Therefore, we have

$$M_m^{[\ell+1]} = \sum_{r=2\ell}^{m-2} (m-r-1)M_r^{[\ell]}.$$

As announced, $M_0^{[0]}$ is consistent with this relation and we can insert these numbers into an array $F=\{f_{n,k}\}_{n,k\in N}$ by setting

$$f_{h,j} = M_{4h+1-j}^{[h]} \quad \text{or} \quad M_m^{[\ell]} = f_{\ell,4\ell-m+1}.$$

The upper part of the array is given in Table A.1 and the previous relation becomes

$$f_{\ell+1,4\ell-m+5} = (m-2\ell-1)f_{\ell,2\ell+1} + \cdots + 1\times f_{\ell,4\ell-m+3}.$$

By setting $4\ell-m+5=k+2$, this relation becomes

$$f_{\ell+1,k+2} = 1\times f_{\ell,k} + 2f_{\ell,k+1} + \cdots + (2(\ell+1)-k)f_{\ell,2\ell+1}.$$

Since for $j>2\ell+1$ we have $f_{\ell,j}=0$, we can extend the sum to

$$f_{\ell+1,k+2} = \sum_{j=0}^{\infty} (j+1)f_{\ell,k+j}.$$

TABLE A.1
The Array F with the Number of (4,2)-Sequences

$\ell\backslash k$	1	2	3	4	5	6	7	8	9
0	1								
1	3	2	1						
2	22	16	10	4	1				
3	211	158	105	52	21	6	1		
4	2306	1752	1198	644	301	116	36	8	1

The relation can be extended to columns 1 and 2, and therefore we have

$$f_{\ell+1,1} = 3f_{\ell,1} + 4f_{\ell,2} + 5f_{\ell,3} + \cdots,$$

$$f_{\ell+1,2} = 2f_{\ell,1} + 3f_{\ell,2} + 4f_{\ell,3} + \cdots,$$

$$f_{\ell+1,3} = 1f_{\ell,1} + 2f_{\ell,2} + 3f_{\ell,3} + \cdots.$$

From these expressions, we immediately deduce that the differences $f_{\ell+1,1} - f_{\ell+1,2}$ and $f_{\ell+1,2} - f_{\ell+1,3}$ are equal and are exactly the row sums $\sum_{j=1}^{\infty} f_{\ell,j}$. These sums are $1, 6, 53, 554, \ldots$, the total number of sequences we are looking for. We now introduce from the literature (see, e.g., [5,8]) some concepts related to Riordan arrays. A *proper Riordan array* is an infinite lower triangular array $D = \{d_{n,k}\}_{n,k \in N}$ for which a sequence $A = \{a_0, a_1, a_2, \ldots\}$ exists such that

$$d_{n+1,k+1} = a_0 d_{n,k} + a_1 d_{n,k+1} + a_2 d_{n,k+2} + \cdots = \sum_{j=0}^{\infty} a_j d_{n,k+j} \qquad \forall n, k \in \mathbf{N}.$$

This sequence, and its generating function $A(t) = \sum_k a_k t^k$, are called the *A-sequence* of the proper Riordan array. Another way to characterize proper Riordan arrays is through two formal power series $D = (d(t), h(t))$, for which we have

$$d_{n,k} = [t^n] d(t)(th(t))^k \qquad \forall n, k \in \mathbf{N}.$$

There exists a simple relation between the two formal power series $h(t)$ and $A(t)$: $h(t) = A(th(t))$. Also, $d(t)$ is simply the generating function for column 0 and is independent of $h(t)$ and $A(t)$. A (*horizontally*) *p-stretched* Riordan array is an infinite lower triangular array $S = \{s_{n,k}\}_{n,k \in N}$ for which a sequence $\hat{A}_p = \{a_0, a_1, a_2, \ldots\}$ exists such that:

$$d_{n+1,k+p} = a_0 d_{n,k} + a_1 d_{n,k+1} + a_2 d_{n,k+2} + \cdots = \sum_{j=0}^{\infty} a_j d_{n,k+j} \qquad \forall n, k \in \mathbf{N}.$$

The discussion above shows that F is a 2-stretched Riordan array having $\{1, 2, 3, \ldots\}$ as A-sequence, or $\hat{A}(t) = (1-t)^{-2}$. In order to study this array, let us proceed as follows. It is known and can be easily proved that if $\hat{A}(t)$ is the A-sequence of a 2-stretched Riordan array, then it can be merged into a proper Riordan array having A-sequence $A(t) = \sqrt{\hat{A}(t)}$, so that in our case we have $A(t) = (1-t)^{-1}$. Let us now define a proper Riordan array $D = \{d_{n,k}\}_{n,k \in N} = (d(t), h(t))$ embedding F and having $\{1, 1, 1, \ldots\}$ as

TABLE A.2
The Proper Riordan Array D Embedding F

$\ell \backslash k$	0	1	2	3	4	5	6	7
0	1							
1	0	**1**						
2	1	1	1					
3	0	**3**	**2**	**1**				
4	6	6	6	3	1			
5	0	**22**	**16**	**10**	**4**	**1**		
6	53	53	53	31	15	5	1	
7	0	**211**	**158**	**105**	**52**	**21**	**6**	**1**

A-sequence. By the relation $h(t) = A(th(t))$, we find $h(t) = C(t) = (1 - \sqrt{1-4t})/2t$, the generating function for the Catalan numbers. In order to find $d(t)$ we perform the following procedure. In Table A.2, we give the upper part of the array and mark in boldface the elements already known from the array F. By applying the rule of the A-sequence from right to left to rows of odd index $2n+1$, we determine all the elements of row $2n$. For example, since $d_{5,5} = a_0 d_{4,4}$ we find $d_{4,4} = 1$; then from $d_{5,4} = a_0 d_{4,3} + a_1 d_{4,4}$ we obtain $d_{4,3} = 3$; and so on. Finally, we determine the elements $d_{2n-1,0}$ by applying the same rule to $d_{2n,1}$. By using the observation above on the differences $f_{\ell+1,1} - f_{\ell+1,2}$ and $f_{\ell+1,2} - f_{\ell+1,3}$, now valid for $d_{2n+1,1} - d_{2n+1,2}$ and $d_{2n+1,2} - d_{2n+1,3}$, we immediately see that

$$d_{2n,0} = d_{2n,1} \quad \text{and} \quad d_{2n+1,0} = 0 \qquad \forall n > 0.$$

We are now in a position to prove the following result:

THEOREM A.1. *The infinite lower triangular array of Table* A.2 *is the proper Riordan array*

$$\left(\frac{2}{2 - tC(t) + tC(-t)}, C(t) \right). \tag{A.1}$$

Proof. We already found that $h(t) = C(t)$, the Catalan function. If we now denote by $tf(t)$ the generating function of column 1, the generating function of column 0 is

$$d(t) = 1 + t\,\frac{f(t) - f(-t)}{2}.$$

On the other hand, by the Riordan array nature of D, we should also have $tf(t) = d(t)tC(t)$ or

$$f(t) = C(t)\left(1 + t\frac{f(t) - f(-t)}{2}\right).$$

Let us set $F(t) = (f(t) - f(-t))/2$; we obviously have $F(-t) = -F(t)$ and, by changing t into $-t$ in the previous relation, we find

$$f(t) = C(t)(1 + tF(t)) \to f(-t) = C(-t)(1 + tF(t)).$$

We subtract these two expressions and obtain

$$2F(t) = C(t) - C(-t) + tF(t)(C(t) - C(-t))$$

and so

$$\begin{aligned} F(t) &= \frac{C(t) - C(-t)}{2 - tC(t) + tC(-t)} \\ &= \frac{1 - 4t^2 - (1 + 2t)\sqrt{1 - 4t} - (1 - 2t)\sqrt{1 + 4t} + \sqrt{1 - 16t^2}}{4t^3}. \end{aligned}$$

This immediately gives the formula for $d(t) = 1 + tF(t)$. ∎

We can now obtain a formula for the total number of sequences with 2ℓ elements: they are $[t^{2l+2}]d(t) = F_{2\ell+1}$. Fortunately, extracting the coefficient of t^n from $F(t)$ is a routine procedure and we find

$$F_n = \frac{3(n+2)}{(n+3)(2n+3)}\binom{2n+4}{n+2} - \frac{4^{(n+1)/2}}{n+2}\binom{n+3}{(n+3)/2} \qquad (n \text{ odd})$$

or in terms of ℓ

$$M_\ell = \frac{3(2\ell+3)}{2(\ell+2)(4\ell+5)}\binom{4\ell+6}{2\ell+3} - \frac{4^{\ell+1}}{2\ell+3}\binom{2\ell+4}{\ell+2} \sim \frac{(3\sqrt{2}-4)16^{\ell+1}}{2\ell\sqrt{\pi\ell}}.$$

Another way to approach the $(4,2)$-case of the tennis ball problem is to define combinatorial objects which are related to the tree of Fig. A.1 as generating trees are related to the tree describing the $(s,1)$-case. Let us consider a tree from the root of which six branches originate; they end at six nodes labeled $(0,0), (0,1), (0,2), (1,1), (1,2), (2,2)$, i.e., all the possible couples from the set $\{0,1,2\}$ having the first element not greater than the second. Then, from each node labeled (k_1,k_2) with $k_1 \leqslant k_2$ we have $\binom{k_1+4}{2}$ branches leading to nodes labeled $(0,0), (0,1), \ldots, (k_1+2, k_1+2)$. The tree so produced can be seen as a generalization of a generating tree with specification:

$$\begin{cases} root: & (0,0), \\ rule: & (k_1,k_2) \rightarrow (0,0)(0,1)\ldots(k_1+2,k_1+2) \qquad k_1 \leqslant k_2. \end{cases}$$

Let us call the tree thus produced 2-*labeled generating tree*; we explicitly observe that the resulting tree is not strictly a generating tree according to the definition usually found in the literature. There is a simple 1–1 correspondence between this tree and the tree shown in Fig. A.1: it is sufficient that, at level ℓ, we let the couple (k_1,k_2) in the 2-labeled generating tree correspond to $(4\ell - k_2 - 1, 4\ell - k_1)$ in the tree of the $(4,2)$-case (observe that the role of the two terms is exchanged). This implies a simple relation between the total sums of the labels in the sequences of the 2-labeled generating tree and those of the $(4,2)$-case tree.

Theorem A.2. *If W_ℓ and $\bar{W}_\ell$ denote the total sum of all the M_ℓ sequences in the* 2-*labeled generating tree and in the* $(4,2)$-*case tree*, *respectively*, *then the following relation holds*:

$$W_\ell + \bar{W}_\ell = M_\ell(4\ell^2 + 3\ell) \sim \frac{2\sqrt{\ell}(3\sqrt{2}-4)16^{\ell+1}}{\sqrt{\pi}}. \tag{A.2}$$

Proof. If (k_1,k_2) is a couple at level j in the 2-labeled generating tree, the corresponding couple in the (4,2)-case tree is $(4j-k_2-1, 4j-k_1)$, and therefore, the contribution to the total sum is $4j-k_2-1+4j-k_1 = 8j-1-(k_1+k_2)$, where k_1+k_2 is the contribution of the couple in the 2-labeled generating tree. Consequently, the sum of the two contributions is $8j-1$ and extended to all the ℓ couples in a sequence becomes $\sum_{j=1}^{\ell}(8j+1) = 4\ell^2+3\ell$. This quantity, multiplied by the total number M_ℓ of sequences, immediately gives the desired result. The approximation comes from the asymptotic value for M_ℓ. ∎

It is rather surprising that the infinite triangle of Table A.3 has an important interpretation for our 2-labeled generating tree:

Theorem A.3. *Let $(r,\cdot)$ be a couple beginning with r in the* 2-*labeled generating tree of Fig.* A.2; *then the number of nodes at level ℓ in the subtree starting from $(r,\cdot)$ is $\mu_r^{[2\ell]} = [t^{2\ell}]d(t)C(t)^{r+2}$, i.e., the 2ℓ coefficient in the $(r+2)$th column of the proper Riordan array of Table* A.2, *apart from a t^{r+2}*.

Proof. Let $f_r(t)$ be the generating function counting the nodes at every level of the subtree starting at a node labeled $(r,\cdot)$; the node produces 1 node

TABLE A.3
The $f_k(t)$'s Coefficients for $k = 0, \ldots, 4$

$f_0(t)$	$f_1(t)$	$f_2(t)$	$f_3(t)$	$f_4(t)$
1	1	1	1	1
6	10	15	21	28
53	105	185	301	462
554	1198	2304	4088	6832
6363	14 506	29 482	55 354	97 957

labeled $(r+2,\cdot)$, 2 nodes labeled $(r+1,\cdot)$, up to $r+3$ nodes with label $(0,\cdot)$. Therefore, $\forall k \geqslant 0$, we have

$$\frac{f_k(t)-1}{t} = (k+3)f_0(t) + (k+2)f_1(t) + \cdots + 2f_{k+1}(t) + f_{k+2}(t) \qquad \text{(A.3)}$$

and obviously $[t^0]f_k(t) = 1$. Table A.3 illustrates some values of the $f_k(t)$'s coefficients for $k = 0, \ldots, 4$: we explicitly observe that (A.3) and the initial conditions determine uniquely these coefficients. Now, we observe that these numbers also appear in some positions of the Riordan array D. In particular, for $n \geqslant 2$, $k \geqslant 0$, we have

$$\begin{aligned} d_{2n+k,k+2} = {} & (k+3)d_{2n-2,2} + (k+2)d_{2n-1,3} + \cdots \\ & + 2d_{2n+k-1,k+3} + d_{2n+k,k+4}. \end{aligned} \qquad \text{(A.4)}$$

This relation can be easily proved by computing the difference $d_{2n+k+1,k+3} - d_{2n+k,k+2}$ and passing to generating functions. In fact, if we set:

$$g^{[k]}(t) = \frac{d(t)C(t)^k - 1 - kt}{t^2}, \qquad h^{[k]}(t) = d(t)\frac{C(t)^k + C(-t)^k}{2},$$

then we obtain the relation

$$\frac{g^{[k+3]}(t) + g^{[k+3]}(-t)}{2} - \frac{g^{[k+2]}(t) + g^{[k+2]}(-t)}{2} = h^{[2]}(t) + \cdots + h^{[k+5]}(t),$$

hence

$$\begin{aligned} & \frac{g^{[k+3]}(t) - g^{[k+2]}(t)}{2} + \frac{g^{[k+3]}(-t) - g^{[k+2]}(-t)}{2} \\ & = \frac{d(t)}{2}\left(C(t)^2 \frac{C(t)^{k+4} - 1}{C(t) - 1} + C(-t)^2 \frac{C(-t)^{k+4} - 1}{C(-t) - 1} \right). \end{aligned}$$

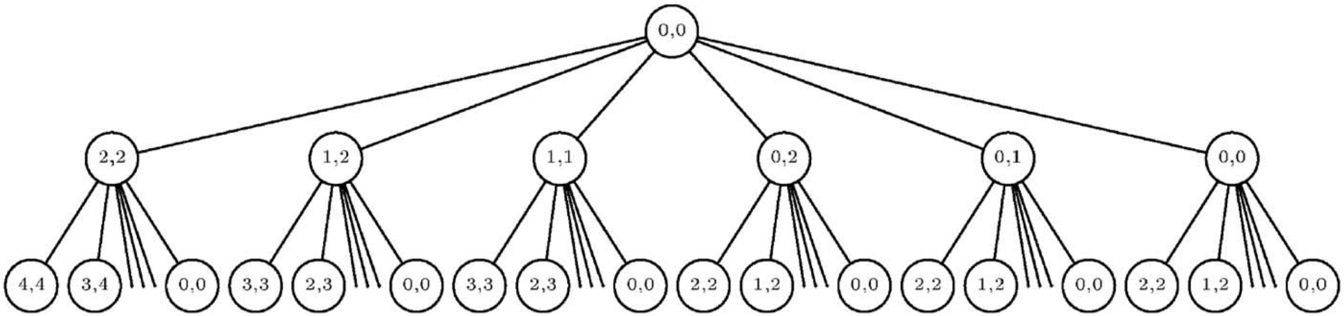

FIG. A.2. The 2-labeled generating tree for the $(4, 2)$-case.

By using the identities $C(t)^2 = (C(t) - 1)/t$ and $C(-t)^2 = (1 - C(-t))/t$ which hold true for the Catalan generating function, we get

$$\frac{d(t)C(t)^{k+4} - 1}{2t} - \frac{d(t)C(-t)^{k+4} - 1}{2t} = d(t)\frac{C(t)^{k+4} - C(-t)^{k+4}}{2t}.$$

Identity (A.4) and the conditions $f_k(0) = 1,\ \forall k \geqslant 0$ imply the statement of the theorem (in fact, the even coefficients in the generating function of column 2, divided by t^2, correspond to the coefficients of $f_0(t)$, the even coefficients in the generating function of column 3, divided by t^3, correspond to $f_1(t)$, etc.). ∎

What we wish to do is to compute the total weight of all the sequences extending up to level ℓ. It is a simple matter to write a Maple procedure generating all the sequences and counting the total sum of their elements; in a few seconds we obtain the results up to level 4, which is sufficient to have some checking values for our developments (see Table A.4). Theorem A.3 gives us a method for counting how many times a couple appearing at level h is counted when we consider all the sequences of length 2ℓ, i.e., arriving to the level $\ell > h$ in the tree. In other terms, $\mu_r^{[2\ell-2h]} = [t^{2\ell-2h}]d(t)C(t)^{r+2}$ represents how many sequences with 2ℓ elements have a couple $(r, \cdot)$ in their positions $2h - 1$ and $2h$. What we are now going to do is to determine the total weight $w_r^{[2h]}$ that the couples $(r, \cdot)$ have at level h; in this way, the product $\mu_r^{[2\ell-2h]} w_r^{[2h]}$ gives us the total weight of couples $(r, \cdot)$ at level h in all the sequences of length 2ℓ. If W_ℓ is this last quantity, the quantity which solves the (4,2)-case of the tennis ball problem, we have

$$W_\ell = \sum_{h=0}^{\ell} \sum_r \mu_r^{[2\ell-2h]} w_r^{[2h]} = \sum_r \sum_{h=0}^{\ell} \mu_r^{[2\ell-2h]} w_r^{[2h]}. \tag{A.5}$$

The internal sum is a convolution and since we know $\mu_r(t) = d(t)C(t)^{r+2}$, we only need to find the generating functions $w_r(t) = \sum_k w_r^{[k]} t^k$. In order to

TABLE A.4
Results Obtained by Maple

Level	No. of sequences	Total weight
1	6	12
2	53	284
3	554	5436
4	6362	96 768

TABLE A.5
The Number of Couples at Level 2

	$(\cdot,0)$	$(\cdot,1)$	$(\cdot,2)$	$(\cdot,3)$	$(\cdot,4)$
$(0,\cdot)$	6	6	6	3	1
$(1,\cdot)$		6	6	3	1
$(2,\cdot)$			6	3	1
$(3,\cdot)$				3	1
$(4,\cdot)$					1

perform this task, we observe again that the elements in row $2h$ of the proper Riordan array D represent how many couples $(\cdot, s)$ are present at level h (irrespective of the value r in the first position). For example, at level 2, we have the couples specified in Table A.5 (this property justifies the Riordan array nature of the triangle D and allows us to compute the values of the row $2h+1$). We can, therefore, compute the total weight of the couples $(r,\cdot)$ by the formula

$$\sum_{k\geqslant r} d_{2h,k}(r+k) = 2r\sum_{k\geqslant r} d_{2h,k} + \sum_{j\geqslant 0} jd_{2h,r+j}.$$

By our construction, $(2r)\sum_{k\geqslant r} d_{2h,k} = 2rd_{2h+1,r+1} = 2r[t^{2h}]d(t)t^r C(t)^{r+1}$. Furthermore, $\sum_{j\geqslant 0} jd_{2h,r+j}$ can be seen as the product between the proper Riordan array D and the column whose generating function is $t^{r+1}/(1-t)^2$. Therefore, we have

$$\begin{aligned}\sum_{j\geqslant 0} jd_{2h,r+j} &= [t^{2h}]D * \frac{t^{r+1}}{(1-t)^2} = [t^{2h}]d(t)\frac{(tC(t))^{r+1}}{(1-tC(t))^2}\\ &= [t^{2h}]d(t)t^{r+1}C(t)^{r+3}. \qquad \text{(A.6)}\end{aligned}$$

Here, we used the well-known relation $C(t) = 1/(1-tC(t))$ valid for the Catalan function $C(t)$. In Fig. A.3 we give the upper part of the triangles involved in this operation. As the result above indicates, what we have obtained is just the proper Riordan array D shifted up by two positions and left by three. The total weight is now given by

$$\begin{aligned}w_r^{[2h]} &= [t^{2h}]d(t)t^{r+1}C(t)^{r+3} + 2r[t^{2h}]d(t)t^r C(t)^{r+1}\\ &= [t^{2h}]d(t)t^r C(t)^{r+1}(tC(t)^2 + 2r), \qquad \text{(A.7)}\end{aligned}$$

or, in terms of generating functions, $w_r(t) = d(t)t^r C(t)^{r+1}(tC(t)^2+2r)$. In Fig. A.4, we show the upper part of the resulting triangle. The final step is to perform sum (A.5). Actually, only the terms of even position are involved in the convolution. Therefore, we should eliminate terms in odd positions, and

$$\begin{pmatrix} 1 \\ 0 & 1 \\ 1 & 1 & 1 \\ 0 & 3 & 2 & 1 \\ 6 & 6 & 6 & 3 & 1 \\ 0 & 22 & 16 & 10 & 41 \\ 53 & 53 & 53 & 31 & 15 & 5 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 & 0 \\ 2 & 1 & 0 \\ 3 & 2 & 1 & 0 \\ 4 & 3 & 2 & 1 & 0 \\ 5 & 4 & 3 & 2 & 1 & 0 \\ 6 & 5 & 4 & 3 & 2 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 & 0 \\ 3 & 1 & 0 \\ 10 & 4 & 1 & 0 \\ 31 & 15 & 5 & 1 & 0 \\ 105 & 52 & 21 & 6 & 1 & 0 \\ 343 & 185 & 80 & 28 & 7 & 1 & 0 \end{pmatrix}$$

FIG. A.3. The product between the triangles involved in (A.6).

$$\begin{pmatrix} 0 \\ 1 & 0 \\ 3 & 1 & 0 \\ 10 & 4 & 1 & 0 \\ 31 & 15 & 5 & 1 & 0 \\ 105 & 52 & 21 & 6 & 1 & 0 \\ 343 & 185 & 80 & 28 & 7 & 1 & 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 & 2 \\ 0 & 4 & 4 \\ 0 & 12 & 12 & 6 \\ 0 & 32 & 40 & 24 & 8 \\ 0 & 106 & 124 & 90 & 40 & 10 \\ 0 & 316 & 920 & 312 & 168 & 60 & 12 \end{pmatrix} =$$

$$\begin{pmatrix} 0 \\ 1 & 2 \\ 3 & 5 & 4 \\ 10 & 16 & 13 & 6 \\ 31 & 47 & 45 & 25 & 8 \\ 105 & 158 & 145 & 96 & 41 & 10 \\ 343 & 501 & 500 & 340 & 175 & 61 & 12 \end{pmatrix}$$

FIG. A.4. The sum of the triangles involved in (A.7).

this is done in a standard way:

$$\hat{\mu}_r(t) = \frac{\mu_r(t) + \mu_r(-t)}{2}, \qquad \hat{w}_r(t) = \frac{w_r(t) + w_r(-t)}{2}.$$

Consequently, for the generating function $W(t)$, we have

$$W(t) = \sum_r \bar{\mu}_r(t)\bar{w}_r(t) = \frac{1}{4}\sum_r (\mu_r(t)w_r(t) + \mu_r(-t)w_r(t) + \mu_r(t)w_r(-t) \\ + \mu_r(-t)w_r(-t)).$$

The four summations are performed quite easily. We only give details for the first one:

$$S_1(t) = \sum_r \mu_r(t)w_r(t) = \sum_r d(t)C(t)^{r+2}d(t)t^rC(t)^{r+1}(tC(t)^2 + 2r)$$

$$= d(t)^2 tC(t)^3 \sum_r (tC(t)^2)^r + 2d(t)^2 C(t)^3 \sum_r (tC(t)^2)^r$$

$$= \frac{td(t)^2 C(t)^5}{1 - tC(t)^2} \left(1 + \frac{2}{1 - tC(t)^2} \right).$$

It is easily observed that $d(-t) = d(t)$, being $d(t)$ an even function, and in a similar way, we get

$$S_2(t) = \sum_r \mu_r(-t) w_r(-t) = -\frac{td(t)^2 C(-t)^5}{1 + tC(-t)^2} \left(1 + \frac{2}{1 + tC(-t)^2} \right),$$

$$S_3(t) = \sum_r \mu_r(t) w_r(-t) = \frac{td(t)^2 C(t)^2 C(-t)^2}{1 - tC(t)C(-t)} \left(C(t) + \frac{2C(-t)}{1 - tC(t)C(-t)} \right),$$

$$S_4(t) = \sum_r \mu_r(-t) w_r(t) = -\frac{td(t)^2 C(t)^2 C(-t)^2}{1 + tC(t)C(-t)} \left(C(-t) + \frac{2C(t)}{1 + tC(t)C(-t)} \right).$$

All these computations were made manually. At this point, since $W(t)$ is a rather complicated function, we can use Maple to perform many tedious tasks. By defining

```
> C := (1 - sqrt(1 - 4*t))/(2*t);
```

we easily find (formula (A.1))

$$d(t) = \frac{2}{1 + \frac{1}{2}\sqrt{1 - 4t} + \frac{1}{2}\sqrt{1 + 4t}}.$$

We also have

$$S_1(t) = \frac{(1 - \sqrt{1 - 4t})^5 (1 - 8t - \sqrt{1 - 4t})}{t^3 (2 + \sqrt{1 - 4t} + \sqrt{1 + 4t})^2 (1 - 4t - \sqrt{1 - 4t})^2},$$

$$S_1(t) = 3t + 20t^2 + 108t^3 + 512t^4 + 2326t^5 + 10\,152t^6 + O(t^7).$$

The expression for $S_3(t)$ is more complicated:

$$S_3(t) = \frac{4(1 - \sqrt{1 - 4t})^2 (1 - \sqrt{1 + 4t})^2 (1 - 4t - (1 + 2t)\sqrt{1 - 4t} - (1 - 6t)\sqrt{1 + 4t} + \sqrt{1 - 16t^2})}{t^3 (2 + \sqrt{1 - 4t} + \sqrt{1 + 4t})^2 (1 + 4t - \sqrt{1 - 4t} - \sqrt{1 + 4t} + \sqrt{1 - 16t^2})^2},$$

$$S_3(t) = 3t + 4t^2 + 34t^3 + 56t^4 + 392t^5 + 720t^6 + O(t^7).$$

Finally, $S_2(t) = S_1(-t)$ and $S_4(t) = S_3(-t)$ and if we define

$$W(t) = \frac{(S_1(t) + S_2(t) + S_3(t) + S_4(t))}{4},$$

the series development of $W(t)$ gives

$$W(t) = 12t^2 + 284t^4 + 5436t^6 + 96768t^8 + 1664184t^{10} + O(t^{11}),$$

which perfectly agrees with the experimental data of Table A.4. If we perform a simple substitution:

$> \texttt{T := subs(t = sqrt(y), W);}$

we obtain a function in which $[y^\ell]T(y)$ is the total sum of all the sequences extending up to the level ℓ. This function can be studied more easily than $W(t)$, because it is not aerated. The presence of $\sqrt{1-16y}$ indicates that the dominating singularity is $y = \frac{1}{16}$. We can ask Maple to expand $T(y)$ into a Taylor series around this singularity and we find a bit surprisingly:

$$T(y) = \frac{32}{(2+\sqrt{2})^2}\frac{1}{1-16y} - \frac{32\sqrt{2}(3+\sqrt{2})}{(2+\sqrt{2})^3}\frac{1}{\sqrt{1-16y}} + O(1).$$

This shows that $T(y)$ has a pole at $y = \frac{1}{16}$. An a posteriori analysis of the denominators in $S_1(t),\ S_2(t),\ S_3(t),\ S_4(t)$ confirms this fact. The previous development implies an asymptotic formula

$$W_\ell \sim \frac{2 \times 16^{\ell+1}}{(2+\sqrt{2})^2}\left(1 - \frac{\sqrt{2}(3+\sqrt{2})}{(2+\sqrt{2})\sqrt{\pi\ell}}\right).$$

To check the validity of this formula, we computed two values:

ℓ	True value	Approx. value	Error
17	$0.613052225 \times 10^{21}$	$0.60751467 \times 10^{21}$	0.91%
47	$0.91656389 \times 10^{57}$	$0.91492571 \times 10^{57}$	0.18%

which shows that the formula is quite accurate. At this point, because of formula (A.2), we have the value of $\bar{W}_l$, and this solves the (4,2)-case of the tennis ball problem:

$$\bar{W}_\ell \sim \frac{2\sqrt{\ell}(3\sqrt{2}-4)16^{\ell+1}}{\sqrt{\pi}} - \frac{2 \times 16^{\ell+1}}{(2+\sqrt{2})^2} = O(\sqrt{\ell}16^{\ell+1}).$$

ACKNOWLEDGMENTS

We thank Lou Shapiro for his encouragement and the referee who suggested the new problem treated in the appendix.

REFERENCES

1. F. R. K. Chung, R. L. Graham, V. E. Hoggat, and M. Kleiman, The number of Baxter permutations, *J. Combin. Theory Ser. A* **24** (1978), 382–394.
2. R. L. Graham, D. E. Knuth, and O. Patashnik, "Concrete Mathematics," Addison–Wesley, Reading, MA, 1989.
3. R. Grimaldi and J. Moser, The Catalan numbers and the tennis ball problem, *Congr. Numer.* **125** (1997), 65–71.
4. C. L. Mallows and L. W. Shapiro, Balls on the lawn, *J. Integer Sequences* **2** (1999).
5. D. Merlini, "Riordan array nell'analisi degli algoritmi," Tesi di Dottorato, Università degli Studi di Firenze, Italy, 1996.
6. D. Merlini and M. C. Verri, Generating trees and proper Riordan Arrays, *Discrete Math.* **218** (2000), 167–183.
7. M. P. Schützenberger, Context-free language and pushdown automata, *Inform. and Control* **6** (1963), 246–264.
8. R. Sprugnoli, Riordan arrays and combinatorial sums, *Discrete Math.* **132** (1994), 267–290.
9. R. Sprugnoli and M. C. Verri, Asymptotics for Lagrange inversion, *Pure Math.* **5** (1994) 79–104.
10. T. Tymoczko and J. Henle, "Sweet Reason: A Field Guide to Modern Logic," Freeman, New York, 1995.
11. J. West, Generating trees and the Catalan and Schroder numbers, *Discrete Math.* **146** (1995), 247–262.
12. J. West, Generating trees and forbidden subsequences, *Discrete Math.* **157** (1996), 363–374.